Personal Information Management

Edited by

William Jones and Jaime Teevan

University of Washington Press
Seattle and London

We dedicate this book to our mothers,
Lorna L. Buckingham-Jones and Connie M. Teevan

Printed in the United States of America
Designed by Ted Cotrotsos
12 11 10 09 08 07 5 4 3 2 1

University of Washington Press
P.O. Box 50096, Seattle, WA 98145 U.S.A.
www.washington.edu/uwpress

Library of Congress Cataloging-in-Publication Data
Personal Information Management / edited by William Jones and Jaime Teevan.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-295-98755-2 (hc : alk. paper)
ISBN-13: 978-0-295-98737-8 (pbk. : alk. paper)
ISBN-10: 0-295-98755-3 (hc : alk. paper)
ISBN-10: 0-295-98737-5 (pbk. : alk. paper)
1. Personal Information Management. I. Jones, William P., 1952 – II. Teevan, Jaime.
HD30.2.P472 2007
650.1—dc22 2007010636

This book is printed on New Leaf Ecobook 50, which is 100 percent recycled, containing 50 percent post-consumer waste, and is processed chlorine free. Ecobook 50 is acid free and meets the minimum requirements of ANSI/NISO Z39~V48~W1992 (R1997) (Permanence of Paper).

6 Save Everything: Supporting Human Memory with a Personal Digital Lifetime Store

Desney Tan, Emma Berry, Mary Czerwinski, Gordon Bell, Jim Gemmell, Steve Hodges, Narinder Kapur, Brian Meyers, Nuria Oliver, George Robertson, and Ken Wood

6.1 Introduction

One of the things that distinguishes human beings from other species is the magnitude to which we manipulate our (largely synthetically created) environments and our technologies in order to augment ourselves physically and mentally. Supporting our individual as well as collective memory has been a particularly important endeavor as we have continued to build upon past experiences and improve our way of life. We are now at a time when each of us is generating and handling more information than ever before. Fortunately, we are now also equipped with technologies that can begin to record, store, summarize, and retrieve all this content. Various governments have recognized the potential of realizing these augmentations and created programs to fund work in the area, as for example in Memories for Life Grand Challenge in the UK, Fitzgibbon and Reiter (2003), and the LifeLog program in the United States (DARPA, n.d.), the work of which has now been distributed into multiple other programs.

In chapter 3 of this book, Jones describes the distinction between information keeping, whereby a user decides what to store, and information organizing, the creation of the structures within which information is stored and retrieved. He reviews strategies and tools that make decisions about keeping information easier for users by having better organization in place from the outset. In this chapter, we build on this work and attempt to completely remove information-keeping decisions by employing what Jones calls a "keep everything" strategy. We explore how we could build personal digital stores that save every bit of information we have touched or record every event we have experienced through our entire lifetime. We believe that this strategy has the potential to support users ranging from the most organized to the least organized (e.g., Alex, Connie, and Brooke). We examine the implications this has on how we organize and retrieve this information.

6.2 Research Overview

The idea of supporting human memory with an all-encompassing personal digital store is not new. In 1945, Vannevar Bush proposed, in his now-iconic article "As We May Think," that "instruments are at hand which, if properly developed, will give man access to and command over the inherited knowledge of the ages" (Bush 1945, p. 101). He described an all-inclusive personal information management (PIM) system he called Memex (from memory extender), which would store all his books, records, communications, and experiences. The user would of course be able to consult the system with exceeding speed and flexibility, and "with one item in its grasp, it snaps instantly to the next that is suggested by the association of thought" (p. 106).

Various aspects of Bush's Memex vision have already been realized. While we do not have the space to review all of this work, we briefly discuss a few important examples. Recording devices that increasingly capture more data, and more *kinds* of data, have been implemented by various wearable computing researchers like Steve Mann, who has also considered some of the related social, artistic, and legal issues (Mann & Niedzviecki 2001). Similarly, the University of Tokyo has developed a system that continuously captures video, along with other sensor data, including GPS, gyroscope, accelerometer, and a brain wave sensor (Hori & Aizawa 2003). Wearable A/V capture systems have even been augmented by interaction with a robot in an exhibition setting (Sumi, Ito, Matsuguchi & Fels 2004).

Extending the use of such recording devices, researchers at Rank Xerox used Active Badges and automatic video recording to automatically generate user diaries (Lamming & Newman 1992) and researchers at AT&T used Active Bats and audio recording for similar purposes (Harter, Hopper, Steggles, et al. 2002). Other researchers have built wearable memory-augmentation devices. For example, the Remembrance Agent provides users with a heads-up display, one-handed chording keyboard, and location awareness, and runs note-taking software that selects old notes to show based on current location, people nearby, and the text of notes being written (Rhodes 2003).

Many storage and implicit query ideas were realized as digital technology as early as the 1960s by Douglas Engelbart, whose hypermedia groupware system supported bookmarks, hyperlinks, recording of email, and a journal (Engelbart, Watson & Norton 1973). Similarly, hypertext visionary Ted Nelson advocated keeping personal recordings of everything and suggested novel computational infrastructure (Nelson 1999). More recently, storage for large personal archives of digital data has been explored by projects such as MIT's Haystack (Adar, Karger & Stein 1999) and Microsoft's Stuff I've Seen (Dumais, Cutrell, Cadiz, et al. 2003). To tie these efforts together, Bell (2001) describes the CyberAll project, aimed at creating a system that records, stores, and allows easy retrieval of a person's entire information for personal and professional use. This project could be considered a predecessor to the efforts to build MyLifeBits, which we describe in the next section.

Since our memories are critical to nearly every activity we perform, we assert that an appropriately designed personal digital store could significantly empower us in our daily lives. For example, in the simplest scenario, the store could be used to *remind* us of things we need to do, or provide information we need to have in various contexts. It could also allow us to quickly *retrieve* digital artifacts such as minutes from a specific meeting or an email message received, perhaps years ago. Since the store records life events as well, it could help us *recall* specific memories such as the names and faces of people we have met, or where we last left our keys. In most current systems, retrieving a document requires remembering its name, deducing keywords that are likely to be in it, or thinking about other properties of the document. Designed correctly, the user would also be able to *recollect* some aspect of the past and use it as a cue for searching. For example, a person may retrieve meeting minutes by remembering something about the event, such as where it was held and who was present.

More complex queries and associations could enable us to *remember* much richer aspects of otherwise forgotten experiences, or to view the experiences from a radically different perspective. For example, a user might look at the amount of time spent on various tasks within the day in order to *reflect* upon and improve their time management skills. Alternatively, they could examine behavioral patterns to monitor medical conditions and improve personal health.

A personal digital store should also promote *reminiscing*, or *reexperiencing* the past for purely social reasons. Reminiscing can be a rewarding experience and an end in itself, supporting the feeling of *reliving* a shared experience as well as providing a basis for storytelling. Sometimes a user might even prefer to let the system choose the memories to be viewed, either because they do not have a specific experience in mind, or because serendipitous reminiscence is often quite pleasurable. When done with others, such activities could allow people to *relate* their memories to others and could help to socially and emotionally connect people.

Our approach to exploring the space of personal digital lifetime stores has been to interactively design and build systems that we can deploy and observe as people use them. We begin this chapter with a broad view of the technologies we have been working on. We first describe our efforts in creating our MyLifeBits infrastructure, as well as SenseCam, a specific recording device that we believe is important for logging life events. We then describe a specific case study in which we examine the effectiveness of these technologies for the therapeutic support of a woman with severe memory loss. This support encompasses many of the activities we have described and provides encouraging evidence for the usefulness of a personal digital store. Finally, we again take a broad view and discuss some of the interesting development and deployment challenges we have experienced. We also consider some of the broader research questions that systems like MyLifeBits inevitably raise with respect to the technical, personal, social, and legal issues associated with storing vast quantities of data about one's life.

6.3 Diving In

6.3.1 *MyLifeBits: Infrastructure for a Personal Digital Store*

MyLifeBits is the infrastructure upon which we base most of our work. MyLifeBits provides a pluggable architecture that allows us to encode, store, access, and manipulate vast arrays of personal information (Gemmell, Bell, Lueder, et al. 2002; Gemmell, Williams, Wood, et al. 2004). This system was initially built to understand how we could digitize a lifetime of legacy content and eliminate paper as a storage medium. As we worked more on MyLifeBits, the original goal of storing files of scanned papers quickly evolved into an exploration of future computing possibilities.

Two of our goals in building this infrastructure were aimed at addressing problems raised in multiple chapters within this book. First, we hoped to unify the various data "islands" created by current application and data-storage boundaries. We believe that partitioning information the way our systems currently do creates artificial restrictions on how flexibly we can interact with information. Second, we wanted to substantially improve the ability to organize, annotate, search, and utilize the vast quantities of content with which we were working. Faced with folders full of digital media with rich metadata tags, we needed a framework to hold and link all of these objects in the arbitrary, Web-like fashion that Bush described.

With these goals in mind, MyLifeBits was built on a SQL Server database that stores content and metadata for a variety of item types. Currently, the MyLifeBits database supports 25 item types, such as contacts, documents, email, events, photographs, music, and video—but arbitrarily many could be added. Each item is stored in a database table containing both shared and type-specific attributes. Items can be linked either implicitly with attributes such as time, or explicitly with typed links such as a "person in photograph" link between a contact and a picture. While many of these links are still manually created, we hope to continue to automate more and more of this process. In this scheme, the traditional directory tree is replaced by more general "collections" in which objects (including a collection) can be filed in more than one parent collection.

Rather than trying to capture and convert small amounts of data from many people, we have focused most of our efforts on creating a single large store containing records of Gordon Bell's life experiences, including a document and activity archive. We now use this store as a vehicle for our research. In addition to continuing conversion of paper documents and physical artifacts, we also populate the database with data from legacy applications such as NTFS files and Outlook email stores. These sources are monitored, and their metadata are integrated along with the full text of each item to enable optimal search. The system also captures items such as Web pages visited, Instant Message chat sessions, telephone conversations, office and meeting audio and video, as well as radio and television program usage. Furthermore, an interface logger records all mouse and keyboard activity. This log can reveal the significance of an item based on frequency or recency of use, or can provide insight into how one spends time with the computer.

To facilitate retrieval and interaction with the data, the core MyLifeBits user interface allows users to query the data and visualize results as a list, variable sized thumbnails, or on a timeline. Users can explore the data by pivoting according to metadata and links. Users can also annotate content with text and voice comments, or assign them to collections. Various plug-ins provide additional functionality to the core system. These currently include a screensaver that displays random photos and video segments and gives the user an opportunity to comment and rate items, simple authoring tools that create side-by-side timelines and HTML-based slide shows with audio, as well as FacetMap, a visual approach that exposes more of the metadata to support more effective data browsing (Smith, McCown & Nelson 2006).

As we gain experience building the infrastructure and utilities, and as we observe more and more people using MyLifeBits with their own data, we continue to iterate on its design and evolve our ideas. In the next section, we describe SenseCam, a wearable system we developed to extend data-capture facilities and allow people to easily record rich life experiences into MyLifeBits.

6.3.2 *SenseCam: Capturing Life*

A personal digital store is only as useful as the information it has available to it, and while capturing artifacts that are created as a result of computer use is a good start, we believe that our everyday experiences are also an extremely rich source of data. Hence we have designed SenseCam, a wearable digital camera that takes photographs passively, without user intervention, while it is being worn (Hodges, Williams, Berry, et al. 2006).

Unlike a regular digital camera or a cameraphone, SenseCam does not have a viewfinder or a display that can be used to frame photos. Instead, it is fitted with a wide-angle (fish-eye) lens that maximizes its field of view. This ensures that nearly everything in view of the wearer is captured by the camera. This is important because a regular wearable camera would likely produce many uninteresting images. SenseCam also contains a number of different electronic sensors. These include light-intensity and light-color sensors, a passive infrared (body heat) detector, a temperature sensor, a multiple-axis accelerometer, and audio level detection. We are also currently working on integrating audio recording and GPS location sensing into SenseCam. These sensors are monitored by the camera's microprocessor, and certain changes in sensor readings can be used to automatically trigger a photograph to be taken. For example, a significant change in light level, or the detection of body heat in front of the camera can cause the camera to take a picture. Alternatively, the user may elect to set SenseCam to operate on a timer, for example taking a picture every thirty seconds.

In our current design, users typically wear the camera on a cord around their neck, although it would also be possible to clip it to pockets or on belts, or to attach it directly to clothing (see Figure 6.3.2). One advantage of using a neck-cord to wear the camera is that it is reasonably stable when being worn—it tends not to move around from left to right when the wearer is walking or sitting—but at the same time it is relatively comfortable to wear and it is easy to put on and take off. Also, when worn around the neck, SenseCam is reasonably close to the wearer's line of sight, and so generates images taken from the wearer's point of

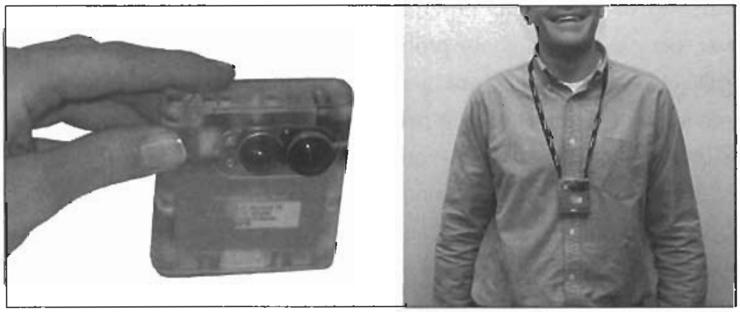

Figure 6.3.2 – The SenseCam v2.3 prototype is shown as a stand-alone and as typically worn by a user. The model pictured here has a clear plastic case that reveals some of the internal components.

view—that is, it creates a "first person" view. Informal observations suggest that this results in images that are more compelling when subsequently replayed.

SenseCam currently takes pictures at VGA resolution (640 x 480 pixels) and stores them as compressed .jpg files on internal flash memory. Surprisingly, most users seem happy with the relatively low-resolution images, suggesting that the time-lapse, first-person viewpoint sequences represent a useful media type that exists somewhere between still images and video. It also points to the fact that these are used as memory supports rather than rich media. Along with the images, SenseCam also stores a log file, which records other sensor data along with their timestamps. Once imported to a PC, files can be stored and manipulated in MyLifeBits, or any other such application.

6.3.3 *Case Study: SenseCam in Support of Severe Memory Loss*

In this section, we describe a specific case study in which we explore the potential of SenseCam and a personal digital store to support a patient with severe memory loss due to brain injury. While this is a single case study, and more work is needed to understand how users adopt such technology in their lives, we believe that the findings from this study are interesting and expose many issues surrounding the development, deployment, and adoption of such technologies.

This work is based on the assertion that psychological intervention and rehabilitation can alleviate some of the debilitating memory and cognitive problems encountered by people with brain injury. It is also based on the assertion that one of the most valuable and effective ways to aid rehabilitation is to use external memory aids to help people to compensate for their memory deficits (Kapur, Glisky & Wilson 2002). Most research on external memory aids aim to improve prospective memory functioning, or memory for future events, such as remembering to keep appointments (Wilson, Evans, Emslie & Malinek 1997). However, people with memory problems following brain damage often experience difficulties with their autobiographical memory, or memory for both remote and recent events from their own personal past. In this case study, we evaluated the use of a personal digital store as an aid to improving autobiographical memory for past events in a patient with amnesia following brain inflammation. We hoped that the patient could use SenseCam images as a pictorial diary to trigger and consolidate her memories. We compared this to a condition in which the same patient used a written diary to aid recall of recent events.

Method

Mrs. B. is a well-educated, 63-year-old woman, and her husband, Mr. B., is a retired businessman. In March 2002 Mrs. B. was admitted to the hospital and diagnosed with limbic encephalitis (inflammation of deep structures of the brain). MRI scans revealed bilateral cell loss in the hippocampus. Mrs. B. would only have partial recall of events that happened in her life after a couple of days, and would typically have no recall at all after about a week. This was true even of significant events. Because of her memory problems, Mrs. B. said she "lacked confidence" in company and was generally anxious in everyday life.

We designed this study to have two conditions: a SenseCam condition and a written diary (control) condition. Medical evidence suggested that Mrs. B. was making a good recovery and that her cognitive functioning would improve over time. Hence, to avoid favoring the SenseCam condition, we decided that it should take place first over a period of three months. The written diary control condition replicated the SenseCam condition as far as possible and took place later over a period of one month. During these time periods, we had Mr. and Mrs. B. use either the SenseCam or a written diary to record events that they particularly wanted Mrs. B. to remember (i.e., beyond everyday, routine events).

In the SenseCam condition, Mr. and Mrs. B. would upload the relevant SenseCam images from an event they wanted to remember to a laptop computer at their earliest convenience. In the written diary condition, Mrs. B. did not wear a SenseCam. Instead, Mr. B. wrote a diary of the events. In both cases, Mr. and Mrs. B. did not look at the images or view the diary together that day. The next day, Mr. B. would ask his wife if she recalled the previous day's events, and noted what she said. Mr. B. would then immediately show Mrs. B. the SenseCam images or written diary of the previous day's events. They viewed and talked about the images or diary entries up to three times that day. Two days later, Mr. B. would again ask his wife what she remembered of the event. Again, he recorded her responses and then showed her the SenseCam images or written diary as before. They repeated the procedure six times, once every two days.

In this way, Mr. B. was able to keep a log of how many times Mrs. B. had viewed the images or diary of a recent event, and her corresponding recall of that event. Mrs. B.'s responses were then graded on a scale of 0 to 100 percent. For each event, Mr. B. documented a number of key points that he felt were important or memorable. For example, if Mr. B. recorded 10 key events, and Mrs. B. remembered 7 of them, she scored 70 percent.

In addition to this short-term testing of recall, we also carried out a longer-term assessment of these autobiographical memories. In the SenseCam condition, Mr. B. showed Mrs. B. all nine of the SenseCam movies that had been created at the end of the three-month period during which the device was used. She then did not view the images for one month, and at the end of that month, Mr. B. tested his wife on her recall of all nine events. He graded her responses in the same way he did before. Two months after this test, Mrs. B. was again presented with images from all nine events, but this time was asked not to view them for a period of two months. She was then tested at the end of these two months. Immediately after this trial, Mrs. B. was again presented with images from all nine events, but this time was asked not to view them for a period of three months. Note that this means that the time interval between the original occurrence of some of the earlier events and eventual testing was as much as 11 months.

In the diary condition, after a month using the diary, Mr. B. showed Mrs. B. the three written diary events that had been created. She then did not view the diary entries for a month, and at the end of this month, Mr. B. tested Mrs. B. on her recall of all three events. He again graded her responses.

Results

Mr. and Mrs. B. recorded nine key events in the SenseCam condition and three key events in the written diary condition. Mrs. B.'s average recall across the events with successive SenseCam/written diary viewings is illustrated in Figure 6.3.3. The results show that there is a significant effect across SenseCam presentations ($x2$ (1) = 62.59, p<0.001). In other words, the more SenseCam images were seen, the greater the percentage of key points about an event were recalled. No such effect was found across the written diary viewings ($x2$ (1) = 0.29, p > 0.6). The results also show that there was an overall difference in average percentage of recall, with SenseCam giving rise to a higher rate of recall than the written diary ($x2$ (1) = 13.10, p < 0.001).

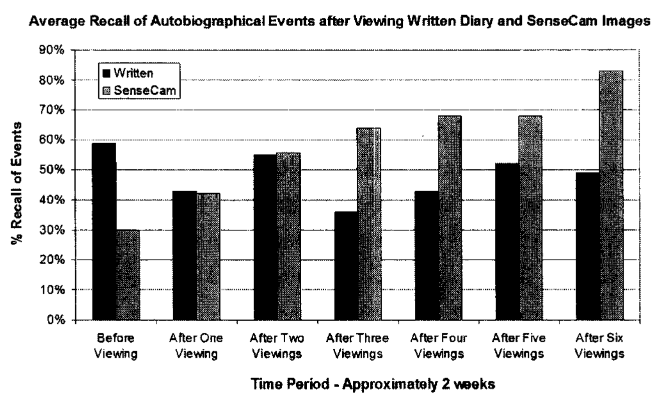

Figure 6.3.3 – The graph represents Mrs. B's average percentage recall of events with successive SenseCam and written diary viewings. The time between viewings was typically two days.

With regard to longer-term effects, Mrs. B.'s recall of events in the SenseCam condition was 80 percent, 67 percent, and 76 percent for one-, two-, and three-month lags respectively. In the written diary condition, she could not recall anything (0 percent) even after one month. No subsequent testing with this technique took place after the one-month interval. This shows that recall with SenseCam remained quite high even after relatively long intervals despite the fact that no additional viewing of the images took place within that interval. This can be contrasted with the written diary, which failed to trigger any recall for Mrs. B. even after a one-month interval.

Discussion

These results show that Mrs. B. demonstrated a consistent improvement in her recent autobiographical memory of events after viewing SenseCam images, which can be contrasted to her results with a written diary. This beneficial effect was maintained in the longer term, many months after the event, and without looking at the SenseCam images for up to three months.

An important point here is that Mrs. B.'s descriptions demonstrated that she remembered the event itself, rather than remembering the SenseCam pictures alone. This was indicated by the fact that Mrs. B. could recall details of the events not captured in the images. In addition, Mr. B. was always quick to point out when his wife was having a true recollective experience as opposed to retelling facts or a narrative that she had learned. Thus, while we would like to collect more evidence, we are encouraged that SenseCam was able to help consolidate memories and re-evoke experiences using the images as triggers. We are also very encouraged by Mrs. B.'s reports that SenseCam reduced her levels of anxiety and increased her confidence, and Mr. B.'s expression of "sheer pleasure" at being able to share experiences with his wife again.

We believe that the positive results seen with SenseCam can be partially attributed to the support it provides for rich visual imagery, which is commonly considered to be associated with autobiographical memory (Brewer 1986, 1988). In fact, brain regions underlying visual memory are an important part of the network that subserves autobiographical memory (Conway 2005; Greenberg & Rubin 2003). SenseCam is different from standard photographic diaries in that it produces a "movie" taken from the point of view of the wearer. When the images are played back, the user is able to see a concise time-lapse movie of their day, similar to the way they experienced it, and these images may be providing a powerful set of stimuli or triggers that cue recall of previously stored memories. Of course, this is a single case study and further research is needed to evaluate the usefulness of SenseCam and personal digital stores in aiding memory and rehabilitation for people with different forms of memory loss.

6.4 Looking Forward

Our attempts to build important components of a Memex-like system in the MyLifeBits platform, and our testing of the SenseCam part of this system with a brain-injured person, raise a number of research challenges both in the short and the long term, and in terms of technical and social issues. In this final section, we discuss some of these challenges as well as potential solutions we have considered. Many of these challenges have also been discussed in varying levels of detail in other chapters. We break the discussion down into subsections dealing with (1) data capture (chapters 4, 5, 9); (2) storage (chapters 4, 5, 8, 9); (3) retrieval, sense-making, and presentation (chapters 4, 8, 10, 14); (4) security, privacy, and access control (chapters 14 through 17); and finally (5) legal and societal issues (chapters 16 and 17).

6.4.1 *Data Capture*

One of the central assumptions of the Memex vision is that the more data one captures, and the more *kinds* of data one captures, the better the system will be. While it may seem tempting to label many artifacts of everyday life as disposable, it is often impossible to predict exactly which items in our lives we might value or need in the future. Hence, the easiest and safest thing is to record as much of it as possible, especially if this requires no effort on the part of the user. The more the system logs, the better the chance of having the memory hook that will help users find what they seek.

In line with the philosophy to record as much as we can, it is essential that we treat anything accessible outside a person's control as transient, and hence create a copy of it in the person's own Memex. For example, while Web-page capture initially struck us as somewhat trivial, we now see this as an essential feature that has changed our everyday behavior. The Internet is constantly morphing, and having a cached personal copy of the particular version viewed is essential. In fact, this is true of much of the information we deal with, even locally.

In practice, of course, some kinds of data are currently more difficult to capture than others. For example, scanning every book that we read is not something we have incorporated into MyLifeBits. The benefits of having this data do not yet outweigh the cost for acquiring appropriate copyright permissions and digitizing content. The reality is that there are many kinds of data that cannot yet be effortlessly and automatically captured.

Hence, we believe that we need to carefully consider the reasons for capturing data and to prioritize various sources and media. For example, we saw in the study with Mrs. B. that the use of SenseCam images provided powerful triggers for recollection. Following this, psychologists have begun to speculate whether there is something special about images, and in particular the kinds of images that SenseCam captures, that may support recollective memory. Studies such as these can help direct and decide where we might best expend our resources with regard to capture. It also points to the fact that there are many unanswered questions as to the data that is the most potent in triggering the kinds of memory we wish to support.

Fortunately, more and more traditional content is being "born digital," reducing the need for expensive conversion. We expect that this will eventually be true of all information, including bills, correspondence, financial statements, news, scholarly articles, music, health data, photos, and more. An added benefit of creating digital content is the opportunity for metadata to be included with little overhead. For example, rather than requiring artificial intelligence to determine the metadata for your dental bills, the software that generates the e-bill could embed the metadata when the content is generated (including that it is a dental bill, whom it is from, for what service, the total due, etc.). Attaching metadata to everything, however, presents a challenge for software to process content and convert it among competing schema. We will discuss the need for a more comprehensive ontology in more detail when we discuss retrieval, sense-making, and presentation.

6.4.2 *Storage*

For decades, storage was a scarce and expensive resource. Today, it is plentiful and affordable. In fact, the storage density a dollar can buy has doubled every 12 months and shows little signs of slowing. If this trend continues, we should see desktop PCs with terabyte storage devices by 2008. When we started this project, we calculated that without audio and video a terabyte would be more than adequate for a lifetime storage of information that a user touches, as it provides more than 1 GByte/month for the duration of an 80-year life. However, changing usage patterns are quickly invalidating this assumption.

While we started thinking we could record things speculatively, storing only digital content we might want to see later, our recent efforts to capture everything have pushed us far beyond what could be stored in a terabyte. Additionally, we have moved beyond storing legacy content—such as paper, photos, and video—into a second phase that includes real-time capture of conversations, meetings, sensor readings, health monitors, and computer activity. In short, while a terabyte might hold a lifetime at 20th-century resolutions and quantities, we speculate that 21st-century users may expect to record their lives more extensively and in higher fidelity—and will drive the market for much greater storage. Even with current trends our ability to generate data will far outpace our ability to store it, and we will necessarily have to consider solutions that employ servers or distributed systems to store our lifetimes of data. Along with this distribution, it is crucial that Memex-like systems be adequately backed up. The failure of one author's hard drive resulted in losing months worth of captured data. Even now, he searches for information expected to be in the archive, only to realize that there is a fairly sizable gap in his personal digital store. Much future work remains on the architecture and reliability of such systems if they are to scale appropriately.

Similarly, as the hardware requirements grow with much more data, we will have to consider scaling the software infrastructure as well. The original instantiation of MyLifeBits utilized only a generic file system with careful naming of files and judicious use of folders and shortcuts. However, as the collection grew the use of files in folders went from unwieldy to overwhelming. Current desktop search tools still work for searching repositories of files and folders. As the amount and types of information grew, we quickly needed more powerful capabilities, such as access by metadata—including written and spoken comments about items and the ability to organize and classify items in multiple ways. The current MyLifeBits uses an SQL-database as the underlying data store for the information, and we imagine that this will continue to evolve as demands get larger and more complex.

6.4.3 *Retrieval, Sense-making, and Presentation*

With recent advances in technology, we are making it easier and easier to create, receive, record, store, and accumulate digital materials. However, it is still extremely difficult to manage and use them in a sensible way, especially as time passes and their immediacy has faded. While some of the issues faced in this research are not new, many of the larger research challenges revolve around coping with the sheer quantity of material that we are now able to record and store.

For example, the strategy of recording and keeping everything often raises concern that clutter may obscure valuable content and add to an individual's personal information management burden, creating distracted attention, information overload, and less-effective searching and browsing (Marshall & Jones 2006). We believe that well-designed technology must hide details and deletions, thus eliminating clutter and the oppressive task of managing it, while still retaining these records for future use. Furthermore, information must be abstracted and displayed in useful and attractive visualizations if people are to be able to interpret and use it appropriately. This must ultimately be linked to an understanding of the particular kinds of tasks people want to carry out with these systems, and the goals they have in mind as they use the data.

When the MyLifeBits project began, there were about 30,000 named items placed in about 1,500 folders. Retrieval was principally by folder location and file name. This quickly turned out to be unwieldy. One possible alternative to this might be to store everything in one large folder and retrieve items using a search engine with knowledge of the folders' content, and with no attention to location. However, full text search is not always enough. Many items require other attributes in order to be found. Unfortunately, with the quantities of information we are dealing with, users are not just unwilling to classify, but are also unable to do it.

We are currently exploring a middle ground that lies between the intricately hand-crafted, and arguably brittle, foldering scheme and no organization. To avoid having to become professional curators constructing our own personal classifications, we have become interested in classification sharing. We are experimenting with hierarchical classifications that will be developed by others to be downloaded by the user, and which contain extra information such as synonyms and descriptions to ease their use. These hierarchical classifications should allow scoping searches in a meaningful way and should lead to systems that scale more gracefully.

But even with convenient classifications and labels ready to apply, we are still asking the user to become a filing clerk—manually annotating every document, email, photo, or conversation. More must be done automatically. The first, easy step is to stop throwing out any potentially useful metadata. Time is probably the most important attribute in our database, yet some photo-editing programs erase the value for date taken. Even capture itself must be more automatic on this scale, so that the user is not forced to interrupt their normal life in order to become their own biographer.

While these are some of the general issues we are tackling, focusing on particular domains and applications is helping us to understand some of the specific technical and design issues. Two examples are (1) visualization to support personal reflection, and (2) monitoring and managing personal health information.

Visualization to Support Personal Reflection

Reporting tools with appropriate visualization form a class of very useful applications supporting reflection. A simple query-based tool can be remarkably insightful and useful—whether it is "how I spend my time" or "count the space used" by different items. In the simplest case, the amount of work on a document, spreadsheet, or Web page can be logged and displayed. Alternatively, reports can track what the user is working on or perhaps even thinking about—for example, by plotting the "budget" or "nominating committee" tasks against time.

Programs that can assist in the creation and visualization of trip diaries and key lifetime event stories will considerably increase use, especially for future viewers who have no familiarity with the content. For example, a fishing trip diary with a timeline, animated maps, and annotations is substantially more valuable to us and our progeny than a collection of unlabelled photos in a labeled folder. Visual autobiographies that summarize a person's life could also be envisioned.

Monitoring and Managing Personal Health Information

Another possible application of these personal lifetime stores is in the area of monitoring and managing one's health (Oliver & Flores-Mangas 2006). This would imply that our system should hold all of a person's health records, associated health financial transactions, and wellness information in order to assist users, caregivers, and health-care providers. Wearable physiological and environmental monitoring is becoming increasingly pervasive, allowing for the continuous, noninvasive collection of many of these signals, such as heart rate, body temperature, and other vital signs. These records could be maintained by an individual and form a baseline indicator of his/her health. Deviations from the average pattern in these metrics could flag potential changes in health.

Important components of such a monitoring system are machine learning and data mining algorithms for finding correlations, trends, and deviations from these trends. These trends would be useful not only in keeping a centralized repository of one's health information, but could also allow the system to provide reminders or actions that would promote better health for the user. The explosion of multisource, heterogeneous, and continuous data in general poses interesting challenges for sense-making utilities.

6.4.4 *Security, Privacy, and Access Control*

As we evolve to record everything for all time, we find significantly more opportunities for research around security, privacy, and access control, as well as ownership and the control of what bits get stored and which get thrown away. If the system is built as a stand-alone computer that stores and retrieves digital records, security is of little concern. Locking a hard drive is not difficult. However, ease of access makes it desirable to connect personal digital memories to a broader virtual network, even the Internet. The network will inevitably connect nearly all of the digital and physical information sources one would like to capture with the digital store. This makes things very convenient, but also very vulnerable, and more work needs to be done in protecting digital stores from malicious attackers.

Security concerns aside, when a user decides to share information with other users, specifying who should have access to the information and in what form imposes a non-trivial burden on the user. Additionally, a particular user may have several different personas within their store. For example, meeting recordings in a workplace setting may be owned by the company and should be partitioned from the rest of the user's social store. When the user moves to a different company, part of their digital memory store might have to be destroyed. Given the volume of data and the diversity of data sources, the choice between private and public could become an onerous one, not just on a daily basis, but also over longer periods of time. We believe that the eventual solution must leverage the nuanced ideas of groups of people and specific contexts within which information should be shared.

Once shared, there are many concerns with the amount of private information that can be inferred from other data. Taken over a long enough period, we expect that everyone will have something recorded that they would prefer to keep private. While there has been a great deal of research in this area (e.g., Ackerman 2000a; Bellotti & Sellen 1993; Cheng, Golubchik & Kay 2004), another chapter in this book concerns itself with these issues specifically, so we will defer to the more thorough treatment of this important subject there.

6.4.5 *Legal and Social Issues*

Many of the issues relating to control, ownership, privacy, security, and modification rights of digital memories are being defined and redefined with new laws in various countries.

It is illegal in some countries to make a copy of anything, including an owner's personal items such as articles, books, and CDs that are protected by copyrights. Similarly, photos of people or objects are protected by copyright such that it is necessary to have permissions for the use of these images. Some of these issues may already be clear from camera, video camera, and phone conversation capture—for example, in many states in the U.S. phone capture is only permitted as long as the party being recorded is informed. Hence, all MyLifeBits dialed or received calls open with a statement that the call is being recorded. Phone calls are really jointly owned, independent of permissions. Hence the ability to share jointly owned content makes the problem even more complex.

Conversion of all information forms to accessible digital forms will create continuing opportunities resulting in needs to access everything, everywhere. In trying to capture more and more data, we may even see a progression from archives on personal computer systems to archives on organizational systems whereby every event is captured throughout a company, government agency, or professional meeting. In this case, privacy and ownership rights will be key impediments and yet safeguards for this vast content.

Organizational relationships are important in our work. While links allow items to be related to one another, we require tools that support doing this across multiple personal stores. In this fashion, one can create family trees, organization charts, or broader social networks. This is especially useful for "contacts," as used by Outlook and other mail systems. Given the incredible mobility of individuals in our personal and professional contact list, it is essential to have a contact for just a company and position, independent of the name of any person who currently holds the position. For example, bankers, brokers, doctors, and other professionals that we maintain relationships with are much more transient than any of us would like. Updating this content is a continual hassle that a store would ideally handle.

6.5 Conclusion

This chapter has introduced the notion of a personal digital lifetime store that could be used for retrieving, recollecting, reminiscing, and reflecting on one's life experiences and information. We described our work on building such a system, including the MyLifeBits platform and the SenseCam data-recording device. We also discussed a specific case study and presented encouraging results for such a SenseCam–based system for supporting a patient with memory deficits. Our goal in writing this chapter was to describe our own personal usage of these technologies to aid and support human memory capabilities, as well as to point toward potential benefits if the technologies are further developed with particular user, corporate, and societal needs in mind. There is much future work that needs to be carried out, both in terms of understanding how useful a personal store of this kind can be, as well as how better to summarize, make sense of, visualize, and share this kind of information. Furthermore, much needs to be done to ensure that the store is of high fidelity, in addition to being safe and secure. Despite the large amount of work that remains, we feel that our early work has begun to demonstrate the real value behind personal digital lifetime stores.

6.6 Acknowledgments

We would like to thank Abigail Sellen for inspiring discussions and insightful feedback on this work; Roger Lueder, Greg Smith, and the rest of the MyLifeBits Development Team; Jim Gray for advice on database aspects of our work; Kentaro Toyama and the World Wide Media Exchange for their assistance with mapping photos; Aleks Aris for making the GPS trails useful as logs resulting in stories; Ed Cutrell and Daniel Robbins for assistance with the graphical user interface for the logger and visualizations; Lyndsay Williams for her instrumental role in CARPE (Continuous Archiving and Retrieval of Personal Experience) and her work with the SenseCam; and everyone else at Microsoft Research who has contributed their ideas, work, and encouragement.